

AD-A169 932

EXTREME VALUES OF BIRTH AND DEATH PROCESSES AND QUEUES  
(U) GEORGIA INST OF TECH ATLANTA R F SERFOZO 25 MAR 86  
AFOSR-TR-86-0384 AFOSR-84-0367

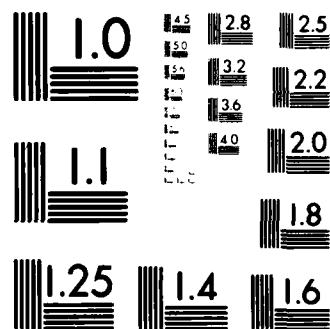
1/1

UNCLASSIFIED

F/G 12/1

NL





MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

6a. NAME OF PERFORMING ORGANIZATION Georgia Institute of Technology		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM	
6c. ADDRESS (City, State and ZIP Code) Atlanta, Georgia 30332			7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC 20332	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR		8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR 84-0367	
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC			10. SOURCE OF FUNDING NOS.	
			PROGRAM ELEMENT NO. 6.1102F	PROJECT NO. 2304
11. TITLE (Include Security Classification) Extreme Values of Birth and Death Processes and Queues			TASK NO. A5	WORK UNIT NO.
12. PERSONAL AUTHOR(S)				
13a. TYPE OF REPORT Preprint		13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) March 25, 1986	
15. PAGE COUNT 26				
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB. GR.	Extreme values, birth and death processes, M/M/s queues, limit theorems. / log document	

EXTREME VALUES OF BIRTH AND DEATH PROCESSES AND QUEUES

by

Richard F. Serfozo

Georgia Institute of Technology

Abstract

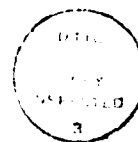
We study the asymptotic behavior of maximum values of birth and death processes over large time intervals. In most cases, the distributions of these maxima, under standard linear normalizations, either do not converge or they converge to a degenerate distribution. However, by allowing the birth and death rates to vary in a certain manner as the time interval increases, we show that the maxima do indeed have three possible limit distributions. Two of these are classical extreme value distributions and the third one is a new distribution. This third distribution is the best one for practical applications. Our results are for transient as well as recurrent birth and death processes and related queues. For transient processes, the focus is on the maxima conditioned that they are finite.

Keywords: Extreme values, birth and death processes, M/M/s queues, limit theorems.

This research was sponsored in part by grant AFOSR 84-0367.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)  
NOTICE OF TRANSMITTAL TO DTIC  
This technical report has been reviewed and is  
approved for public release IAW AFR 190-12.  
Distribution is unlimited.

ANTHONY J. KESTER  
Chief, Technical Information Division



Approved for public release;  
distribution unlimited.

## 1. Introduction

When modeling the dynamics of a parameter of a system by a stochastic process, the questions one addresses depend on the nature of the parameter. In some instances, the extreme values of the parameter rather than its usual values may be of paramount interest. In a manufacturing plant, for example, a typical parameter is the queue length of parts waiting to be processed at a work station. Small to moderate values of the queue may indicate that the system is operating successfully and the queue fluctuations are unimportant. On the other hand, large queues may call for extraordinary measures such as allocation of auxiliary storage space, employee overtime, or rescheduling of production. A natural question is: What is the probability that the queue will exceed a specific critical value in a certain time period? Extreme value questions like this are the topic of this paper. More specifically, our focus is on characterizing the asymptotic behavior of the maxima of birth and death processes and related queues.

The gist of our study is illustrated by the following results for the  $M/M/s$  queue. Consider such a queueing process in which customers arrive to  $s$  servers according to a Poisson process with rate  $\lambda$ , and the independent, exponentially distributed service times have mean  $\mu^{-1}$ . Let  $M_n$  denote the maximum queue length in the time interval  $[0, T_n]$ , where  $T_n$  is the  $n$ th time the system becomes empty. Our interest is in finding norming constants  $a_n, b_n > 0$  and a non-degenerate distribution  $G$  such that

$$(1.1) \quad \lim_{n \rightarrow \infty} P((M_n - a_n)/b_n \leq x) = G(x),$$

for each continuity point of  $G$ . When such  $a_n, b_n, G$  exist, we say that

$M_n$  has the limit distribution  $G$ . Otherwise, we say that  $M_n$  does not have a limit distribution. As usual, we consider only linear normalizations.

We can write  $M_n = \max\{Y_1, \dots, Y_n\}$ , where  $Y_k$  is the maximum of the queue in the interval  $[T_{k-1}, T_k]$ . Since the queuing process is Markovian, then  $Y_1, Y_2, \dots$  are independent identically distributed random variables. From the classical extreme value theory for independent identically distributed variables (see for instance Galambos (1978) or Leadbetter et al. (1983)), we know that the possible limit distributions for  $M_n$  are only  $\exp(-x^{-\gamma})$ ,  $x > 0$ , or  $\exp(-e^{-x})$ ,  $x \in \mathbb{R}$ . One consequence is as follows; this is a special case of Theorem 2.5.

THEOREM 1.1. If the queueing process is null recurrent ( $\lambda = s\mu$ ), then

$$\lim_{n \rightarrow \infty} P(M_n / (n(\lambda/\mu)^{s!}) < x) = e^{-x^{-1}}, \quad x > 0.$$

Cohen (1969) proved an analogue of this for the M/G/1 and G/M/1 queues; related studies on extreme values of queue lengths and waiting times are Heyde (1971) and Iglehart (1972).

Another classical result for discrete random variables is that the convergence (1.1) can take place only if

$$\lim_{m \rightarrow \infty} P(Y_1 = m) / P(Y_1 > m-1) = 0.$$

Consequently, we have the following anomaly.

THEOREM 1.2. If the queueing process is positive recurrent ( $\lambda < s\mu$ ), then  $M_n$  does not have a limit distribution.

This non-convergence theorem is surprising, especially in light of Theorem 1.1, since positive recurrent processes generally have nicer properties than null recurrent ones. Cohen (1969) and Anderson (1970)

give insights on typical  $\liminf$ 's and  $\limsup$ 's for the distribution of  $(M_n - a_n)/b_n$ .

Our point of departure is to establish the convergence (1.1) for queues and birth and death processes in spite of the non-convergence described in Theorem 1.2 and its generalization, Theorem 2.3. Our approach is to allow the birth and death parameters  $(\lambda, \mu, s$  for the  $M/M/s$  queue) to vary with  $n$  when considering the convergence of  $M_n$ . Such parameter variations in limit theorems are not uncommon. A classic example is that if  $S_n$  is a binomial random variable with parameters  $n, p$  and if  $p = p_n$  varies with  $n$  such that  $np_n \rightarrow \lambda > 0$  as  $n \rightarrow \infty$ , then the distribution of  $S_n$  converges weakly to a Poisson distribution with mean  $\lambda$ .

Here is an example of our major results. Suppose  $M_n$  is the maximum, as above, of an  $M/M/s$  queue, where the arrival rate  $\lambda = \lambda(n)$  and service rate  $\mu = \mu(n)$  depend on  $n$ . Let  $\rho_n = \lambda(n)/(s\mu(n))$ .

THEOREM 1.3. Suppose that  $\rho_n < 1$  for each  $n$  and that  $\rho_n \rightarrow 1$ . The possible limit distributions for  $M_n$  are  $G_0(x) = \exp(-x^{-1})$ ,  $x > 0$ ;  $G_\infty(x) = \exp(-e^{-x})$ ,  $x \in \mathbb{R}$ ; and

$$(1.2) \quad G_c(x) = \exp(-c/(e^x - 1)), \quad x > 0, \quad \text{for } 0 < c < \infty.$$

The  $M_n$  has the limit distribution  $G_c$ , where  $0 < c < \infty$ , if and only if  $n(1 - \rho_n)/s! \rightarrow c$ . Appropriate norming constants  $a_n, b_n$  are as in Theorem 3.1.

Note that  $G_0$  and  $G_\infty$  are classical extreme value distributions. The third distribution (1.2) has not appeared in the literature before. Numerical work has shown that this new distribution is the best one for practical approximations. Namely, for the standard  $M/M/s$  queue with traffic intensity  $\rho = \lambda/s\mu$ ,

$$P(-M_n \log \rho < x) \cong \exp(c_n / (e^k - 1)), \quad x > 0,$$

where  $c_n = n(1-\rho)$ . This approximation is good for  $n > 15$  and for all  $\rho < 1$ .

This completes our introduction. Here is what lies ahead: Section 2 consists of preliminaries, including classical convergence and non-convergence theorems for extreme values of birth and death processes; Section 3 contains our main results for recurrent birth and death processes; Section 4 contains analogous results for transient processes; and Section 5 gives applications to M/M/s and related queues.

## 2. Preliminaries

We shall consider a continuous-time birth and death process on the nonnegative integers with birth rates  $\lambda_0, \lambda_1, \lambda_2, \dots$  and death rates  $\mu_0 = 0, \mu_1, \mu_2, \dots$ . This is a Markov process that evolves as follows: Upon entering state  $k$ , the process remains there for an exponentially distributed time with mean  $(\lambda_k + \mu_k)^{-1}$ , and then it moves to state  $k+1$  or  $k-1$  according to the respective probabilities  $\lambda_k / (\lambda_k + \mu_k)$  and  $\mu_k / (\lambda_k + \mu_k)$ . We assume for now that the process is positive recurrent or null recurrent - transient processes are discussed in Section 4. We also assume, for convenience, that the process at time zero begins in state zero. Let  $M_n$  denote the maximum value of the process in the time interval  $[0, T_n]$ , where  $T_n$  is the time of the  $n$ th visit of the process to state zero. The  $M_n$  and  $T_n$  are finite valued since the process is recurrent. We shall study the asymptotic behavior of the distribution of  $M_n$  as  $n \rightarrow \infty$ .

We can write  $M_n = \max\{Y_1, \dots, Y_n\}$ , where  $Y_k$  is the maximum of the birth and death process in the time interval  $[T_{k-1}, T_k]$ , here  $T_0 = 0$ . Since the process is Markovian, the random variables  $Y_1, Y_2, \dots$  are



independent and identically distributed. Consequently,

$$P(M_n \leq x) = P(Y_1 \leq x, \dots, Y_n \leq x) = F(x)^n,$$

where  $F$  is the distribution of the  $Y_k$ 's. To obtain an expression for  $F$ , note that the successive states of the birth and death process form a simple discrete-time random walk that moves from state  $k$  to state  $k+1$  or  $k-1$  according to the respective probabilities  $\lambda_k/(\lambda_k + \mu_k)$  and  $\mu_k/(\lambda_k + \mu_k)$ . Then clearly  $F(x)$  is the probability that, starting from state 1, the random walk reaches state 0 before it exceeds  $x$ . Thus, from Section I.12 of Chung (1967), we know that

$$(2.1) \quad F(x) = 1 - \left( \sum_{k=0}^x r_k \right)^{-1}$$

where  $r_0 = 1$  and  $r_k = (\mu_1 \dots \mu_k) / (\lambda_1 \dots \lambda_k)$ ,  $k \geq 1$ . Furthermore, the birth and death process is recurrent when  $\sum_{k=0}^{\infty} r_k = \infty$  and is transient when  $\sum_{k=0}^{\infty} r_k < \infty$ .

Here are some asymptotic properties of the ratio  $\lambda_k/\mu_k$ , depending on whether the birth and death process is transient or recurrent; we let

$$(2.2) \quad \rho = \liminf_{k \rightarrow \infty} \lambda_k/\mu_k \quad \text{and} \quad \bar{\rho} = \limsup_{k \rightarrow \infty} \lambda_k/\mu_k.$$

LEMMA 2.1. If  $\sum_{k=0}^{\infty} r_k < \infty$ , then  $\lambda_k/\mu_k > 1$  for an infinite number of  $k$ 's, and  $\rho > 1$ . If  $\sum_{k=0}^{\infty} r_k = \infty$ , then  $\bar{\rho} \leq 1$ .

Proof. Suppose  $\sum_{k=0}^{\infty} r_k < \infty$ . For any  $N \geq 1$ , we have

$$r_{N-1} \sum_{k=N}^{\infty} (\mu_N \dots \mu_k) / (\lambda_N \dots \lambda_k) = \sum_{k=N}^{\infty} r_k < \infty.$$

Thus  $\lambda_k/\mu_k$  must be  $> 1$  for an infinite number of  $k$ 's; otherwise, the first sum would be infinite. To prove  $\rho > 1$ , fix  $\epsilon > 0$  and let  $N$  be such

that  $\lambda_k/\mu_k > \rho - \varepsilon$ ,  $k \geq N$ . Then

$$\infty > \sum_{k=0}^{\infty} r_k = \sum_{k=0}^{N-1} r_k + r_N \sum_{m=0}^{\infty} (\rho - \varepsilon)^{-m}.$$

Consequently,  $(\rho - \varepsilon)^{-1} < 1$  for each  $\varepsilon > 0$ , and hence  $\rho > 1$ . The second assertion is proved similarly.

Keep in mind that, for now, we are considering only recurrent birth and death processes. Our interest is in the weak convergence of the distribution

$$(2.3) \quad P((M_n - a_n)/b_n \leq x) = F(a_n + b_n x)^n = [1 - (1 - F(a_n + b_n x))]^n,$$

where  $a_n$  and  $b_n > 0$  are constants. It is well known that, for any  $\gamma_n \in \mathbb{R}$  and  $-\infty < \gamma < \infty$ , the convergence  $(1 + \gamma_n)^n \rightarrow e^\gamma$  is equivalent to  $n\gamma_n \rightarrow \gamma$ .

This property applied to (2.3) translates into the following known result (cf Corollary 1.3.1 of Galambos (1978) or Theorem 1.5.1 of Leadbetter et al. (1983)). Here  $0 \leq g(x) < \infty$ .

CONVERGENCE CRITERION 2.2.  $P((M_n - a_n)/b_n \leq x) \rightarrow e^{-g(x)}$  as  $n \rightarrow \infty$  if and only if  $n(1 - F(a_n + b_n x)) \rightarrow g(x)$  as  $n \rightarrow \infty$ . Note that this criterion is also true when  $F$  varies with  $n$ .

We begin by characterizing when the maximum  $M_n$  does not have a limit distribution and when it might have one. Here we use the notation (2.2)

$$\text{and } \alpha_{km} = \prod_{\ell=k+1}^m \lambda_\ell / \mu_\ell.$$

THEOREM 2.3. (i) The maximum  $M_n$  does not have a limit distribution if and only if

$$(2.4) \quad \liminf_{n \rightarrow \infty} \sum_{k=0}^{m-1} \alpha_{km} < \infty.$$

(ii) Inequality (2.4) holds when  $\sum_{k=0}^{\infty} \limsup_{m \rightarrow \infty} \alpha_{km} < \infty$  or when  $\bar{\rho} < 1$ .

(iii) Inequality (2.4) does not hold when  $\sum_{k=0}^{\infty} \liminf_{m \rightarrow \infty} \alpha_{km} = \infty$  or when  $\rho > 1$ . The condition  $\rho > 1$  is equivalent to  $\lim_{k \rightarrow \infty} \lambda_k / \mu_k = 1$ .

(iv) If  $\alpha_k = \lim_{m \rightarrow \infty} \alpha_{km}$  exists for each  $k$ , then inequality (2.4) holds if and only if  $\sum_{k=0}^{\infty} \alpha_k < \infty$ .

Proof. (i) We know, from Theorem 1.7.13 of Leadbetter *et al.* (1983), that there exist  $x_n$  and  $0 < \tau < \infty$  such that  $n(1 - F(x_n)) \rightarrow \tau$  as  $n \rightarrow \infty$  if and only if

$$(F(m) - F(m-1))/(1 - F(m-1)) \rightarrow 0 \quad \text{as } m \rightarrow \infty.$$

The latter condition, in light of (2.1), is

$$r_m / \sum_{k=0}^{m-1} r_k = \left( \sum_{k=0}^{m-1} \alpha_{km} \right)^{-1} \rightarrow 0 \quad \text{as } m \rightarrow \infty.$$

Then, by Criterion 2.2, the  $M_n$  does not have a limit distribution if and only if the last convergence does not hold, which is equivalent to (2.4).

(ii) If the sum  $\sum_{k=0}^{\infty} \limsup_{m \rightarrow \infty} \alpha_{km}$  is finite, then by Fatou's lemma this finite sum is an upper bound for  $\limsup_{m \rightarrow \infty} \sum_{k=0}^{m-1} \alpha_{km}$  and hence (2.4) holds.

Now suppose  $\bar{\rho} < 1$ . Then for any  $c$  in the interval  $(\bar{\rho}, 1)$ , there is a number  $N$  such that  $\lambda_k / \mu_k < c$  for  $k > N$ . It follows that, for  $m > N$ ,

$$\begin{aligned} (2.5) \quad \sum_{k=0}^m \alpha_{km} &= r_m^{-1} \sum_{k=0}^N r_k + \sum_{k=N+1}^m \alpha_{km} \\ &< r_N^{-1} c^{m-N} \sum_{k=0}^N r_k + \sum_{\ell=0}^{m-N} c^{\ell} \\ &\rightarrow 1/(1-c) \quad \text{as } m \rightarrow \infty. \end{aligned}$$

Thus (2.4) holds.

(iii) By Fatou's lemma, the sum  $\sum_{k=0}^{\infty} \liminf_{m \rightarrow \infty} \alpha_{km}$  is a lower bound for the left side of (2.4). Thus, if this sum is infinite, then (2.4) does not

hold. Now suppose  $\rho > 1$ . Then for any  $c < 1$ , there is an  $N$  such that  $\lambda_k/\mu_k > c$  for  $k > N$ . The display (2.5) clearly holds with the inequality reversed and so

$$\liminf_{m \rightarrow \infty} \sum_{k=0}^m \alpha_{km} > 1/(1-c).$$

Letting  $c \rightarrow 1$  implies that (2.4) does not hold. To prove that  $\rho > 1$  if and only if  $\lim_{k \rightarrow \infty} \lambda_k/\mu_k = 1$ , we need only show that  $\rho > 1$  implies  $\rho = \bar{\rho} = 1$ . But this is true because, by Lemma 2.1, we know that  $\rho < \bar{\rho} < 1$ .

(iv) This part is a consequence of parts (ii) and (iii).

The preceding theorem yields the negative conclusion that the existence of a limit distribution for  $M_n$  is the exception rather than the rule: There is no limit distribution for a typical positive recurrent process with  $\limsup_{k \rightarrow \infty} \lambda_k/\mu_k < 1$ , but one might exist for an atypical process with  $\lim_{k \rightarrow \infty} \lambda_k/\mu_k = 1$  (such as a null recurrent process). For those instances when there might be a limit distribution for  $M_n$ , we have the following properties from the classical extreme value theory.

PROPERTIES OF  $M_n$  2.4. (a) The possible limit distributions for  $M_n$  are only  $\exp(-x^{-Y})$ ,  $x > 0$ , or  $\exp(-e^{-x})$ ,  $x \in \mathbb{R}$ .

(b) The first of these distributions is the limit if and only if

$$(2.6) \quad (1 - F(tx))/(1 - F(t)) = \frac{\sum_{k=0}^t r_k / \sum_{k=0}^{tx} r_k}{\sum_{k=0}^t r_k / \sum_{k=0}^{tx} r_k} \rightarrow x^Y \text{ as } x \rightarrow \infty.$$

Appropriate norming constants are  $a_n = 0$  and  $b_n = \min\{m: \sum_{k=0}^m r_k > n\}$ .

(c) The second distribution in (a) is the limit distribution for  $M_n$  if and only if there is a positive function  $g(t)$  such that

$$(2.7) \quad (1 - F(t + xg(t))) / (1 - F(t)) = \frac{\sum_{k=0}^t r_k / \sum_{k=0}^{t+xg(t)} r_k}{\rightarrow c^{-x} \text{ as } t \rightarrow \infty.}$$

In this case, one can choose

$$g(t) = \frac{t}{\sum_{k=0}^t r_k} \frac{\infty}{\sum_{k=t}^{\infty} (\sum_{m=0}^k r_m)^{-1}},$$

and  $a_n = \min\{m: \sum_{k=0}^m r_k > n\}$  and  $b_n = g(a_n)$ .

A special case of (b) is as follows. This applies, for instance, to a null recurrent process with  $\lambda_k = \mu_k$ ,  $k > s$ , for some  $s$  (such as the M/M/s queue in Theorem 1.1).

THEOREM 2.5. If  $\sum_{k=0}^{\infty} |1 - \lambda_k / \mu_k| < \infty$ , then

$$(2.8) \quad \lim_{n \rightarrow \infty} P(M_n / nb \leq x) = e^{-x^{-1}}, \quad x > 0$$

where  $b = \prod_{k=1}^{\infty} \lambda_k / \mu_k$ .

Proof. A basic property of infinite products of real numbers is that

$\prod_{k=1}^{\infty} (1 - a_k)$  exists when  $\sum_{k=1}^{\infty} |a_k| < \infty$ . In light of this, the hypothesis

implies the existence of the limit  $b$ . Now

$$\lim_{k \rightarrow \infty} r_k = \lim_{k \rightarrow \infty} \prod_{\ell=1}^k \mu_{\ell} / \lambda_{\ell} = b^{-1}.$$

Consequently,  $n^{-1} \sum_{k=0}^n r_k \rightarrow b^{-1}$ , and so

$$n(1 - F(nbx)) = [bx(nbx)^{-1} \sum_{k=0}^{nbx} r_k]^{-1} \rightarrow x^{-1} \text{ as } n \rightarrow \infty.$$

This convergence and Criterion 2.2 yield (2.8).

### 3. Main Results

We saw above that the maximum  $M_n$  does not have a limit distribution for a wide class of birth and death processes, including those in which

$\limsup_{k \rightarrow \infty} \lambda_k / \mu_k < 1$ . However,  $M_n$  might have a limit distribution when  $\lim_{k \rightarrow \infty} \lambda_k / \mu_k = 1$ . These negative and slightly positive findings prompted us to explore the convergence of  $M_n$  when the parameters  $\lambda_k = \lambda_{nk}$  and  $\mu_k = \mu_{nk}$  vary with  $n$  such that  $\lambda_{nk} / \mu_{nk}$  is nearly 1 for large  $n$  and  $k$ . This is the basis of our following results.

Consider a sequence of recurrent birth and death processes indexed by  $n = 1, 2, \dots$ , where the  $n$ th process has the respective birth and death rates  $\lambda_{nk}$  and  $\mu_{nk}$  when in state  $k = 0, 1, \dots$ . For the  $n$ th process, let  $M_n$  denote its maximum up to the time of its  $n$ th return to state zero. This  $M_n$  has the same meaning as the one in Section 2, but here its defining parameters  $\lambda_{nk}, \mu_{nk}$  vary with  $n$  as well as with  $k$ . That is,  $M_n$  is the maximum of  $n$  independent random variables with the common distribution

$$F_n(x) = 1 - \left( \sum_{k=0}^x r_{nk} \right)^{-1}$$

where  $r_{n0} = 0$  and  $r_{nk} = (\mu_{n1} \dots \mu_{nk}) / (\lambda_{n1} \dots \lambda_{nk})$ ,  $k > 1$ .

We shall assume that, for each  $n$ , there is a positive number  $\rho_n < 1$  and a positive integer  $s_n$  such that

$$(3.1) \quad \lambda_{nk} / \mu_{nk} = \rho_n \quad \text{for } k > s_n,$$

and

$$(3.2) \quad \lim_{n \rightarrow \infty} n^{-1} \sum_{k=0}^{s_n} r_{nk} = 0.$$

The parameter  $s_n$  may be bounded, unbounded, or independent of  $n$ , such as  $s_n = 1$ . Assumption (3.1), with  $\rho_n < 1$ , implies that  $\sum_{k=0}^{\infty} r_{nk} > \sum_{k=s_n}^{\infty} \rho_n^{-k}$

$= \infty$ , which ensures that the  $n$ th birth and death process is recurrent.

Assumption (3.1) is satisfied automatically for M/M/s queues: the  $s_n$  is the number of servers. Assumption (3.2) holds when  $s_n$  and  $r_{nk}$  are bounded or when  $r_{nk} \leq B_n$  for  $k \leq s_n$  and  $s_n B_n/n \rightarrow 0$ .

We shall show that when  $\rho_n \rightarrow 1$ , the possible limit distributions for  $(M_n - a_n)/b_n$  are as follows. Here  $c = \lim_{n \rightarrow \infty} c_n$  and

$$(3.3) \quad c_n = n(1 - \rho_n)r_{ns_n}^{-1} = n(1 - \rho_n) \prod_{k=1}^{s_n} \lambda_{nk} / \mu_{nk}.$$

Case	Limit Distribution	Norming Constants
$c = 0$	$G_0(x) = \exp(-x^{-1}), x > 0,$	$a_n = s_n - 1 \quad b_n = n \prod_{k=1}^{s_n} \lambda_{nk} / \mu_{nk}$
$0 < c < \infty$	$G_c(x) = \exp(-c/(e^x - 1))$	$a_n = s_n - 1 \quad b_n = -1/\log \rho_n$
$c = \infty$	$G_\infty(x) = \exp(-e^{-x}), x \in \mathbb{R},$	$a_n = s_n - 1 - \log c_n / \log \rho_n$ $b_n = -1/\log \rho_n.$

An easy check shows that the distributions  $G_c$  ( $0 \leq c \leq \infty$ ) are of distinct type:  $G_c$  and  $G_{c'}$  are of the same type ( $G_c(x) = G_{c'}(a + bx)$  for some  $a, b$ ) if and only if  $c = c'$ .

The following result gives sufficient conditions for the existence of limit distributions for  $M_n$ .

**THEOREM 3.1** Suppose (3.1) and (3.2) hold and  $\rho_n \rightarrow 1$ . If  $c_n \rightarrow c$  as  $n \rightarrow \infty$ , where  $0 \leq c \leq \infty$ , then

$$(3.4) \quad \lim_{n \rightarrow \infty} P((M_n - a_n)/b_n \leq x) = G_c(x), \quad x \in \mathbb{R},$$

where the  $G_c, a_n, b_n$  corresponding to the limit  $c$  are displayed above.

**Proof.** By the Convergence Criterion 2.2, it suffices to show that  $c_n \rightarrow c$  implies

$$(3.5) \quad \lim_{n \rightarrow \infty} n(1 - F_n(a_n + b_n x)) = x^{-1} \quad \text{when } c = 0$$

$$= c/(e^x - 1) \quad \text{when } 0 < c < \infty$$

$$= e^{-x} \quad \text{when } c = \infty.$$

To this end, let  $m_n(x)$  denote the integer part of  $a_n + b_n x$ . Using (3.2), we have

$$\begin{aligned} (3.6) \quad n(1 - F_n(a_n + b_n x)) &= (n^{-1} \sum_{k=0}^{m_n(x)} r_{nk})^{-1} \\ &= (n^{-1} \sum_{k=0}^{s_n-1} r_{nk} + n^{-1} r_{ns_n} \sum_{k=s_n}^{m_n(x)} \rho_n^{-(k-s_n)})^{-1} \\ &= (o(1) + z_n(x))^{-1}, \end{aligned}$$

where

$$\begin{aligned} z_n(x) &= (\rho_n^{s_n - m_n(x)} - \rho_n) / c_n \quad \text{when } \rho_n < 1 \\ &= (m_n(x) - s_n) / (nr_{ns_n}^{-1}) \quad \text{when } \rho_n = 1. \end{aligned}$$

Then to establish (3.5), it suffices to show that  $z_n(x)^{-1}$  converges to the values on the right side of (3.5). Three cases present themselves.

Case 1:  $c_n \rightarrow c = 0$ . Here

$$m_n(x) = a_n + b_n x + o(1) = s_n - 1 + xnr_{ns_n}^{-1} + o(1).$$

First consider the special situation in which  $\rho_n = 1$  for each  $n$ . Using  $n^{-1}r_{ns_n} \rightarrow 0$ , we have

$$\begin{aligned} z_n(x)^{-1} &= nr_{ns_n}^{-1} / (m_n(x) - s_n) \\ &= (x + o(1))^{-1} \rightarrow x^{-1} \text{ as } n \rightarrow \infty. \end{aligned}$$

Next consider the general situation in which  $\rho_n \rightarrow 1$ . Because of the preceding, we may assume that  $\rho_n < 1$  for each  $n$ .

Using

$$\log \rho_n = -(1 - \rho_n) + o(1 - \rho_n) \text{ and } e^u = 1 + u + o(u)$$

as  $u \rightarrow 0$ , it follows that



$$\begin{aligned}
\frac{s_n - m_n(x)}{\rho_n} - \rho_n &= \exp[(s_n - m_n(x)) \log \rho_n] - \rho_n \\
&= \exp[xc_n + O(1 - \rho_n)] - \rho_n \\
&= 1 + xc_n + O(1 - \rho_n) + o(c_n) - \rho_n \\
&= xc_n + O(1 - \rho_n) + o(c_n).
\end{aligned}$$

Substituting this in the expression for  $z_n(x)$ , and using  $(1 - \rho_n)/c_n = n^{-1}r_{ns_n} \rightarrow 0$ , we have the desired convergence

$$z_n(x)^{-1} = [x + O(1 - \rho_n)/c_n + o(1)]^{-1} \rightarrow x^{-1} \quad \text{as } n \rightarrow \infty.$$

Case 2:  $c_n \rightarrow c$  and  $0 < c < \infty$ . Here

$$m_n(x) = s_n - 1 - x/\log \rho_n + O(1).$$

Using  $\log \rho_n = o(1)$ , we have

$$\begin{aligned}
z_n(x)^{-1} &= c_n / (\exp[(s_n - m_n(x)) \log \rho_n] - \rho_n) \\
&= c_n / (\exp[x + o(1)] - \rho_n) \\
&\rightarrow c / (e^x - 1) \quad \text{as } n \rightarrow \infty.
\end{aligned}$$

Case 3:  $c_n \rightarrow c = \infty$ . Here

$$m_n(x) = s_n - 1 - x/\log \rho_n + \log c_n / \log \rho_n + O(1).$$

Using  $\log \rho_n = o(1)$ , we have

$$\frac{s_n - m_n(x)}{\rho_n} = \exp[(s_n - m_n(x)) \log \rho_n] = c_n e^{x+o(1)}.$$

Thus

$$z_n(x)^{-1} = (e^{x+o(1)} - \rho_n/c_n)^{-1} \rightarrow e^{-x} \quad \text{as } n \rightarrow \infty.$$

This completes the proof.

Theorem 3.1 says that the convergence of  $c_n$  is sufficient for  $M_n$  to have a limit distribution when  $\rho_n \rightarrow 1$ . The next result says that the convergence of  $c_n$  is necessary as well, and that  $M_n$  has no limit

distribution other than those above.

**THEOREM 3.2.** Suppose (3.1) and (3.2) hold and  $\rho_n \rightarrow 1$ . The possible limit distributions for  $M_n$  are  $G_c$ ,  $0 < c < \infty$ ; and  $M_n$  has the limit distribution  $G_c$  if and only if  $c_n \rightarrow c$  as  $n \rightarrow \infty$ .

Proof. Suppose  $M_n$  has the limit distribution  $H$ . Since  $[0, \infty]$  is a closed set in the extended real line, there are positive integers  $n_k \rightarrow \infty$  and  $c$  in  $[0, \infty]$  such that  $c_{n_k} \rightarrow c$  as  $k \rightarrow \infty$ . Then by Theorem 3.1, we know that  $M_{n_k}$  has the limit distribution  $G_c$ . Moreover,  $M_{n_k}$  also has the limit distribution  $H$ . From Khintchine's theorem on convergence to types of distributions (see for instance Theorem 1.2.3 of Leadbetter et al. (1983)), it follows that  $H$  and  $G$  are of the same type. This proves that any limit distribution of  $M_n$  must be one of the distributions  $G_c$ ,  $0 < c < \infty$ .

We now prove that  $c_n \rightarrow c$  is necessary and sufficient for  $M_n$  to have a limit distribution. The sufficiency follows from Theorem 3.1. To prove the necessity, suppose that  $M_n$  has the limit distribution  $G_c$ . Let  $c_{n_k}$  be any convergent subsequence of  $c_n$  and let  $c' = \lim_{k \rightarrow \infty} c_{n_k}$ . Arguing as in the last paragraph, it follows that  $G_{c'}$ , as well as  $G_c$  is a limit distribution of  $M_n$  and that  $G_{c'}$  and  $G_c$  are of the same type. Consequently,  $c' = c$ . Thus, we have shown that any convergent subsequence of  $c_n$  must converge to  $c$  and hence  $c_n \rightarrow c$ .

**APPROXIMATION OF  $P(M_n \leq x)$  FOR PRACTICAL APPLICATIONS 3.3.** Consider the maximum  $M_n$  of a birth and death process with rates  $\lambda_k, \mu_k$  (without the artificial dependence on  $n$ ) that satisfy  $\lambda_k/\mu_k = \rho$ ,  $k \geq s$ , for some  $0 < \rho < 1$  and  $s \geq 1$ . Theorems 3.1 and 3.2 yield the approximation

$$(3.7) \quad P((M_n - s + 1)(-\log \rho) \leq x) \cong \exp(c_n/(e^x - 1)), \quad x > 0,$$

where  $c_n = n(1-\rho) \prod_{k=1}^s \lambda_k / \mu_k$  and  $n$  is large.

There are analogous approximations for  $P(M_n \leq x)$  by  $G_0$  when  $c_n$  is small and by  $G_\infty$  when  $c_n$  is large. However, these are not as good as (3.7), which is superior for any  $c_n$ . This is because  $G_0$  and  $G_\infty$  are only theoretical limits for the two "unobtainable" values of  $c_n$  in  $[0, \infty]$ ; they are not functions of the actual  $c_n$  as the right side of (3.7) is. For the case when the process is null recurrent ( $\rho=1$ ), Theorem 2.4 yields the classical approximation  $P(M_n/nb \leq x) \cong G_0(x)$ , where  $b = \prod_{k=1}^s \lambda_k / \mu_k$ . We were pleasantly surprised that the approximation (3.7) is accurate even when  $\rho$  is not near one. This is apparently because  $\rho$  appears on the right as well as left side of (3.7). For the M/M/1 queue, we found from numerical computations that the difference between the two sides of (3.7) is below 0.018 when  $n = 15$  and below 0.01 when  $n = 20$ , for any  $\rho$  in  $(0, 1)$ .

PROPERTIES OF THE LIMIT DISTRIBUTIONS 3.4. Let  $X_c$  denote a random variable with distribution  $G_c$ ,  $0 < c < \infty$ , and let  $Y$  denote an exponentially distributed random variable with unit mean. Standard change-of-variable computations show that

$$X_0 \stackrel{D}{=} 1/Y, \quad X_\infty \stackrel{D}{=} -\log Y, \quad X_c \stackrel{D}{=} \log(Y + c) - \log Y, \quad 0 < c < \infty,$$

where these are equalities in distribution. Solving the equations for  $Y$  and using obvious substitutions, we also have

$$X_0 \stackrel{D}{=} \exp(-X_\infty) \stackrel{D}{=} c / (\exp(X_c) - 1)$$

$$X_\infty \stackrel{D}{=} \log X_0 \stackrel{D}{=} \log((\exp(X_c) - 1)/c)$$

$$X_c \stackrel{D}{=} \log(1 + cX_0) \stackrel{D}{=} \log(1 + c\exp(-X_\infty)).$$

The  $G_\infty$  can be viewed as the limit of  $G_c$  as  $c \rightarrow \infty$  in that

$G_c(x + \log c) \rightarrow G_\infty(x)$  as  $c \rightarrow \infty$  (i.e.  $X_c - \log c \rightarrow X_\infty$ ). It is known that

$$EX_\infty = E(-\log Y) = - \int_0^\infty e^{-y} \log y \, dy = \gamma,$$

where  $\gamma = .5772\dots$  is the Euler-Mascheroni constant  $\gamma = \lim_{m \rightarrow \infty} (\sum_{n=1}^m 1/n - \log m)$ , and

$$E(e^{\alpha X_\infty}) = E(Y^{-\alpha}) = \Gamma(1-\alpha), \text{ for } \alpha < 1,$$

where  $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} \, du$ . In comparison, for  $0 < c < \infty$ , we have

$$\begin{aligned} EX_c &= E(\log(c + Y) - \log Y) \\ &= e^c \int_0^\infty e^{-u} \log u \, du + \gamma. \end{aligned}$$

Hence

$$EX_c = \gamma + e^c(\gamma_c - \gamma),$$

where  $\gamma_c = - \int_0^c e^{-u} \log u \, du$ , which can be interpreted as the

Euler-Mascheroni constant on  $[0, c]$  (recall that  $\gamma_\infty = \gamma$ ). The  $\gamma_c$  can be computed by numerical integration; it is a positive continuous function in  $c$  that increases on  $[0, 1]$  and decreases on  $[1, \infty]$ . Furthermore, using the binomial expansion, we have

$$E(\exp(\alpha X_c)) = E(1 + c/Y)^\alpha = \sum_{k=1}^\infty \binom{\alpha}{k} c^{\alpha-k} \Gamma(1+k-\alpha), \text{ for } \alpha < 1.$$

We end this section with further insights into the irregular behavior of  $M_n$ . Can  $M_n$  have a limit distribution when  $\rho_n$  does not converge to one? Can  $M_n$  have a discrete limit distribution? The following results show that the answer to each of these questions is yes.

**PROPOSITION 3.5.** Suppose the birth and death processes satisfy (3.1), (3.2) and  $\rho_n \rightarrow \rho < 1$ . Let  $a_n$  be a sequence of integers. Then the

distribution of  $M_n - a_n$  converges weakly to a nondegenerate limit  $H$  as  $n \rightarrow \infty$  if and only if

$$(3.8) \quad n^{-1} r_{ns_n} \rho_n^{s_n - a_n} \rightarrow \alpha > 0 \quad \text{as } n \rightarrow \infty.$$

In this case,

$$H(x) = \exp(-\alpha^{-1}(1-\rho)\rho^{[x]}), \quad x \in \mathbb{R},$$

where  $[x]$  denotes the integer part of  $x$  ( $H$  is concentrated on the integers).

Proof. With no loss in generality, we may assume that  $\rho_n < 1$  for each  $n$ .

By (3.6) with  $m_n(x) = a_n + [x]$ , we have

$$n(1 - F_n(a_n + x)) = (o(1) + n^{-1} r_{ns_n} (\rho_n^{s_n - a_n - [x]} - \rho_n) / (1 - \rho_n))^{-1}.$$

Recall that  $n^{-1} r_{ns_n} \rightarrow 0$  because of (3.2). Then clearly

$n(1 - F_n(a_n + x)) \rightarrow \alpha^{-1}(1-\rho)\rho^{[x]}$  if and only if (3.8) holds. Hence, the assertions follow by Criterion 2.2.

Proposition 3.5 may not be too useful for applications since (3.8), apparently, is rarely satisfied. Indeed, we know from Theorem 2.3 (i) that there do not exist  $a_n$  that satisfy (3.8) when  $\lambda_{nk}$ ,  $\mu_{nk}$ ,  $s_n$  and  $\rho_n$  are independent of  $n$ . A subtle variation in these quantities as  $n \rightarrow \infty$  is therefore needed for the existence of  $a_n$  that satisfy (3.8). Here is such an instance.

EXAMPLE 3.6. Consider the special case in which  $\lambda_{nk}/\mu_{nk} = n^{-1}/[\log n]$ ,

$k > s$ , and  $\prod_{\ell=1}^s \lambda_{n\ell}/\mu_{n\ell} \rightarrow \gamma > 0$  as  $n \rightarrow \infty$ . Then (3.1) and (3.2) hold and

$\rho_n = \exp(-\log n/[\log n]) \rightarrow e^{-1}$ . Let  $a_n = [\log n]$ . Then

$$n^{-1} r_{ns} \rho_n^{s-a_n} = r_{ns} \rho_n^s + \gamma^{-1} e^{-s} \quad \text{as } n \rightarrow \infty.$$

Hence Proposition 3.5 yields

$$\lim_{n \rightarrow \infty} P(M_n - a_n \leq x) = \exp(-\beta e^{-[x]}), \quad x \in \mathbb{R},$$

where  $\beta = \gamma e^s (1 - e^{-1})$ . This limit distribution is a discrete version of  $G_\infty(x) = \exp(-e^{-x})$ .

#### 4. Extreme Values of Transient Processes

Consider a birth and death process, as in Section 2, with rates

$\lambda_k, \mu_k$ . Assume that the process is transient, that is, the sum  $B = \sum_{k=0}^{\infty} r_k$

is finite. Let  $M_n$  denote the maximum of the process up to the time  $T_n$  of its  $n$ th return to state 0. Because the process is transient, the  $M_n$  and  $T_n$  may be infinite: from (2.1) we know that

$$P(M_1 \leq x) = 1 - \left( \sum_{k=0}^x r_k \right)^{-1}, \quad \text{and so } P(M_1 < \infty) = 1 - 1/B < 1. \quad \text{Of interest,}$$

therefore, is the asymptotic behavior of  $M_n$  conditioned on  $M_n < \infty$ .

Accordingly, we now consider the convergence of the conditional distribution

$$(4.1) \quad P((M_n - a_n)/b_n \leq x \mid M_n < \infty) = H(a_n + b_n x)^n,$$

where

$$(4.2) \quad H(x) = P(M_1 \leq x \mid M_1 < \infty) = (1 - \left( \sum_{k=0}^x r_k \right)^{-1}) / (1 - 1/B).$$

Similar to the terminology above, we say that  $M_n$  conditioned on  $M_n < \infty$  has a limit distribution or doesn't have one according to whether or not the distribution (4.1) converges weakly to a nondegenerate

distribution. The following result is analogous to Theorem 2.3. Here we

use  $\rho = \liminf_{k \rightarrow \infty} \lambda_k / \mu_k$ ,  $\bar{\rho} = \limsup_{k \rightarrow \infty} \lambda_k / \mu_k$ , and  $\beta_{mk} = \prod_{\ell=1}^k \mu_{m+\ell} / \lambda_{m+\ell}$ .

THEOREM 4.1. (i) The  $M_n$  conditioned on  $M_n < \infty$  does not have a limit distribution if and only if

$$(4.3) \quad \liminf_{m \rightarrow \infty} \sum_{k=m}^{\infty} \beta_{mk} < \infty.$$

(ii) Inequality (4.3) holds when  $\sum_{k=0}^{\infty} \limsup_{m \rightarrow \infty} \beta_{mk} < \infty$  or when  $\rho > 1$ .

(iii) Inequality (4.3) does not hold when  $\sum_{k=0}^{\infty} \liminf_{m \rightarrow \infty} \beta_{mk} = \infty$  or when  $\bar{\rho} < 1$ . The condition  $\bar{\rho} < 1$  is equivalent to  $\lim_{k \rightarrow \infty} \lambda_k / \mu_k = 1$ .

(iv) If  $\beta_k = \lim_{m \rightarrow \infty} \beta_{mk}$  exists for each  $k$ , then inequality (4.3) holds if and only if  $\sum_{k=0}^{\infty} \beta_k < \infty$ .

Proof. From (4.2) and a little algebra, we get

$$(4.4) \quad (H(x) - H(x-1)) / (1 - H(x-1)) = \left( \sum_{k=0}^{\infty} r_k / \sum_{k=0}^m r_k \right) \left( \sum_{k=m}^{\infty} \beta_{mk} \right)^{-1}.$$

Then arguing as in the proof of Theorem 2.3 (i), it follows that  $M_n$  conditioned on  $M_n < \infty$  does not have a limit distribution if and only if expression (4.4) does not converge to zero, which is equivalent to (4.3) (the first term on the right of (4.4) converges to one).

Part (iii) follows since, by Fatou's lemma and the product form of  $\beta_{mk}$ , we have

$$\liminf_{m \rightarrow \infty} \sum_{k=m}^{\infty} \beta_{mk} > \sum_{k=0}^{\infty} \liminf_{m \rightarrow \infty} \beta_{mk} > \sum_{k=0}^{\infty} b^{-k},$$

and the second assertion in (iii) is proved the same way its analogue in

Theorem 2.3 (iii) was. Part (ii) follows by a similar argument, and part (iv) is a consequence of (ii) and (iii).

Theorem 4.1 says that for a typical transient process with  $\liminf_{k \rightarrow \infty} \lambda_k / \mu_k > 1$ , there does not exist a limit distribution for  $M_n$  conditioned on  $M_n < \infty$ , but there might be a limit for a process with  $\lim_{k \rightarrow \infty} \lambda_k / \mu_k = 1$ . When it is possible for the limit to exist, then the asymptotic behavior of  $M_n$  conditioned on  $M_n < \infty$  is analogous to Properties 2.4. Here we have a further simplification.

REMARK 4.2. The distribution (4.1) has the same limiting behavior as the distribution  $\tilde{H}(a_n + b_n x)^n$ , where  $\tilde{H}(x) = B^{-1} \sum_{k=0}^x r_k$ . This follows since

$$1 - H(x) = (1 - \tilde{H}(x)) / ((1 - B^{-1}) \sum_{k=0}^x r_k),$$

and so ratios of the form  $(1 - H(x_n)) / (1 - H(y_n))$ , like (2.6) and (2.7), have the same limiting behavior as  $(1 - \tilde{H}(x_n)) / (1 - \tilde{H}(y_n))$  when  $x_n, y_n \rightarrow \infty$ .

The preceding observations lead to the study of  $M_n$  when the ratio  $\lambda_k / \mu_k$  depends on  $n$  and is nearly unity for large  $n$ . Accordingly, consider the maximum  $M_n$ , as in Section 3, for a sequence of birth and death processes with rates  $\lambda_{nk}, \mu_{nk}$  that depend on  $n$  as well as  $k$ . Then

$$P(M_n \leq x \mid M_n < \infty) = H_n(x)^n$$

where

$$H_n(x) = (1 - (\sum_{k=0}^x r_{nk})^{-1}) / (1 - 1/B_n),$$

and  $r_{n0} = 1$ ,



$$r_{nk} = (\mu_{n1} \dots \mu_{nk}) / (\lambda_{n1} \dots \lambda_{nk}), \quad B_n = \sum_{k=0}^{\infty} r_{nk}.$$

Assume that (3.1) and (3.2) hold and that  $\rho_n > 1$  for each  $n$ . Then each birth and death process is transient since  $B_n$  is finite:

$$(4.5) \quad B_n = r_{ns_n} \sum_{m=0}^{\infty} \rho_n^{-m} = r_{ns_n} / (1 - 1/\rho_n).$$

The following result, analogous to Theorems 3.1 and 3.2 combined, says that when  $\rho_n \rightarrow 1$ , the possible limit distributions for  $M_n$  conditioned on  $M_n < \infty$  are  $G_c$ ,  $0 < c < \infty$ . Here we let

$$c_n = n(1 - 1/\rho_n) r_{ns_n}^{-1} \quad c = \lim_{n \rightarrow \infty} c_n$$

and

$$\begin{aligned} a_n &= s_n - 1 & b_n &= nr_{ns_n}^{-1} & \text{when } c &= 0 \\ a_n &= s_n - 1 & b_n &= 1/\log \rho_n & \text{when } 0 < c < \infty \\ a_n &= s_n - 1 - \log(1/c_n - 1/n)/\log \rho & b_n &= 1/\log \rho_n & \text{when } c &= \infty. \end{aligned}$$

This notation is that of Section 3 with  $\rho_n$  replaced by  $1/\rho_n$  and the last  $a_n$  changed slightly.

**THEOREM 4.3.** Suppose (3.1) and (3.2) hold,  $\rho_n > 1$  for each  $n$ , and  $\rho_n \rightarrow 1$ . Then  $M_n$  conditioned on  $M_n < \infty$  has a limit distribution if and only if  $c_n \rightarrow c$  where  $0 < c < \infty$ . In this case

$$(4.6) \quad \lim_{n \rightarrow \infty} P(M_n - a_n)/b_n \leq x \mid M_n < \infty = G_c(x), \quad x \in \mathbb{R},$$

where  $a_n, b_n$  are defined above.

**Proof.** We will prove that  $c_n \rightarrow c$  implies (4.6). Then the rest of the assertion will follow by the argument we used in the proof of Theorem 3.2. Similar to (4.5), we can write, for  $m \geq s_n$ ,

$$(4.7) \quad \sum_{k=0}^m r_{nk} = r_{ns_n} \sum_{\ell=0}^{m-s_n} \rho_n^{-\ell} = r_{ns_n} (1 - \rho_n^{-m+s_n-1}) / (1 - 1/\rho_n).$$

Let  $m_n(x)$  denote the integer part of  $a_n + b_n x$ . Then using (4.5) and (4.7), we have

$$\begin{aligned}
 (4.8) \quad n(1 - H_n(a_n + b_n x)) &= n(B_n / \sum_{k=0}^{m_n(x)} r_{nk} - 1) / (B_n - 1) \\
 &= c_n ((1 - \rho_n^{-m_n(x) + s_n - 1})^{-1} - 1) / (1 - c_n/n) \\
 &= 1 / ((1/c_n - 1/n)(\rho_n^{m_n(x) - s_n + 1} - 1)).
 \end{aligned}$$

Suppose  $c_n \rightarrow c$ , where  $0 < c < \infty$ . Then by Cases 1 and 2 in the proof of Theorem 3.1, with  $\rho_n$  replaced by  $\rho_n^{-1}$ , it follows that

$$\begin{aligned}
 (4.9) \quad \lim_{n \rightarrow \infty} n(1 - H_n(a_n + b_n x)) &= \lim_{n \rightarrow \infty} c_n / (\rho_n^{m_n(x) - s_n - 1} - 1) \\
 &= x^{-1} \quad \text{when } c = 0 \\
 &= c / (e^x - 1) \quad \text{when } 0 < c < \infty.
 \end{aligned}$$

Next, suppose  $c_n \rightarrow \infty$ . Here

$$m_n(x) = s_n - 1 - \log(1/c_n - 1/n) / \log \rho_n + x / \log \rho_n + o(1).$$

Using this in (4.8), we have

$$\begin{aligned}
 (4.10) \quad \lim_{n \rightarrow \infty} n(1 - H_n(a_n + b_n x)) &= \lim_{n \rightarrow \infty} 1 / ((1/c_n - 1/n) \rho_n^{m_n(x) - s_n + 1} - 1) \\
 &= \lim_{n \rightarrow \infty} \exp(-(m_n(x) - s_n + 1) \log \rho_n - \log(1/c_n - 1/n)) \\
 &= e^{-x}.
 \end{aligned}$$

Thus, (4.9), (4.10) and Criterion 2.2 yield (4.6).

REMARK 4.4. The analogue of Approximation 3.3 for the birth and death process with  $\lambda_k / \mu_k = \rho$ ,  $k \geq s$ , and  $\rho > 1$  is as follows

$$(4.11) \quad P((M_n - s + 1) \log \rho_n \leq x \mid M_n < \infty) \cong \exp(c_n / (e^x - 1)), \quad x > 0,$$

where  $c_n = n(1 - 1/\rho) \prod_{k=1}^s \lambda_k / \mu_k$ .

## 5. Extreme Values of Queues

We now apply the preceding results to the M/M/s and related queues. The M/M/s queueing process described in Section 1 is a birth and death process with birth rate  $\lambda_k = \lambda$  (the Poisson arrival rate of customers) and death rate  $\mu_k = \mu \min\{k, s\}$  (the rate at which  $k$  customers depart from the  $s$  servers).

For our first result, we suppose  $M_n$  is the maximum of this M/M/s queue up to the  $n$ th time the system becomes empty. The limiting behavior of  $M_n$  depends on the queue's traffic intensity  $\rho = \lambda/s\mu$ . The queueing process is positive recurrent when  $\rho < 1$ , null recurrent when  $\rho = 1$ , and transient when  $\rho > 1$ . The following is an immediate consequence of Theorems 2.3, 2.5 and 4.2.

COROLLARY 5.1. If  $\rho < 1$ , then  $M_n$  does not have a limit distribution. If  $\rho = 1$ , then

$$(5.1) \quad \lim_{n \rightarrow \infty} P(m_n/nb \leq x) = e^{-x^{-1}}, \quad x > 0,$$

where  $b = (\lambda/\mu)^{s!}$ . If  $\rho > 1$ , then  $M_n$  conditioned on  $M_n < \infty$  does not have a limit distribution.

For the next result, we suppose that  $M_n$ , as in Sections 3 and 4, is the maximum for an M/M/ $s_n$  queue with arrival rate  $\lambda_{nk} = \lambda(n)$ , service rate  $\mu_{nk} = \mu(n)\min\{k, s_n\}$ , and number of servers  $s_n$ . The traffic intensity of the queue is  $\rho_n = \lambda(n)/(s_n\mu(n))$ . Clearly  $\lambda_{nk}/\mu_{nk} = \rho_n$ , for  $k > s_n$ . We will use

$$r_{ns_n} = \prod_{\ell=1}^{s_n} \mu_{n\ell}/\lambda_{n\ell} = s_n! (\mu(n)/\lambda(n))^{s_n}.$$

COROLLARY 5.2. Suppose  $n^{-1}r_{ns_n} \rightarrow 0$  and  $\rho_n \rightarrow 1$ .

- (i) If  $\rho_n < 1$  for each  $n$ , then  $M_n$  has a limit distribution if and only if  $n(1 - \rho_n)r_{ns_n}^{-1} \rightarrow c$ , where  $0 < c < \infty$ . In this case, the limit distribution is  $G_c$  and appropriate norming constants are as in Theorem 3.1.
- (ii) If  $\rho_n > 1$  for each  $n$ , then  $M_n$  conditioned on  $M_n < \infty$  has a limit distribution if and only if  $n(1 - 1/\rho_n)r_{ns_n}^{-1} \rightarrow c$ , where  $0 < c < \infty$ . In this case, the limit distribution is  $G_c$  and appropriate norming constants are as in Theorem 4.3.

Proof. The two assertions are special cases of Theorems 3.1, 3.2 and Theorem 4.3, respectively. Note that condition (3.1) is satisfied, and so is (3.2) since

$$n^{-1} \sum_{k=0}^{s_n} r_{nk} < 2n^{-1} r_{ns_n} \rightarrow 0.$$

REMARKS 5.3. (a) The number of customers in an  $M/M/\infty$  queueing system over time is a birth and death process with rates  $\lambda_k = \lambda$  (the Poisson arrival rate of customers) and  $\mu_k = k\mu$  (where  $\mu$  is the service rate of each of the infinite servers). The traffic intensity is  $\rho = \lambda/\mu$ . The first and third assertions of Corollary 5.1 also hold for this queue, but there are apparently no analogues of (5.1) or Corollary 5.2.

b) Consider a service system in which the number of customers in the system over time is a birth and death process with rates  $\lambda_k, \mu_k$  that represent customer arrival and departure rates when  $k$  customers are present. We assume that  $\lambda_k/\mu_k = \rho$  for  $k > s$ , where  $s$  is a specific state, but we place no other restriction on the rates. We call this an  $M/M/GR-s$  queueing process, where  $GR$  stands for general rates. General rates are used for modeling such phenomena as balking and reneging of customers; non-standard service disciplines; dynamically changing rates

under a control policy that minimizes the system cost; and simultaneous customer processing, where  $\mu_k$  is the total workrate when  $k$  customers are present and customer  $i$  receives  $p_i$  of the workrate ( $p_1 + \dots + p_m = 1$ ).

Corollaries 5.1 and 5.2 readily extend to M/M/GR-s queues.

(c) The approximations (3.7) and (4.11) apply to the M/M/s and M/M/GR-s queues.

## References

- Anderson, C.W. (1970). Extreme value theory for a class of discrete distribution with applications to some stochastic processes. *J. Appl. Prob.* 7, 99-113.
- Chung, K.L. (1967). *Markov Chains With Stationary Transition Probabilities*. Second ed. Springer-Verlog.
- Cohen, J.W. (1982). *The Single Server Queue*. North-Holland.
- Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. Wiley.
- Heyde, C. (1971). On the growth of the maximum queue length in a stable queue. *Operations Res.* 19, 447-452.
- Iglehart, D.L. (1972). Extreme values in the GI/G/1 queue. *Ann. Math. Statist.* 43, 627-635.
- Leadbetter, M.R., Lindgren, G. and R. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag.

END

DTIC

8-86